

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-263224

(43) 公開日 平成8年(1996)10月11日

(51) Int.Cl.⁶

G 0 6 F 3/06

識別記号

5 4 0

庁内整理番号

F I

G 0 6 F 3/06

技術表示箇所

5 4 0

審査請求 未請求 請求項の数41 O L (全 20 頁)

(21) 出願番号 特願平8-25038

(22) 出願日 平成8年(1996)2月13日

(31) 優先権主張番号 3 9 6 0 4 6

(32) 優先日 1995年2月28日

(33) 優先権主張国 米国 (U S)

(71) 出願人 390009531

インターナショナル・ビジネス・マシーンズ・コーポレーション

INTERNATIONAL BUSINESS MACHINES CORPORATION

アメリカ合衆国10504、ニューヨーク州
アーモンク (番地なし)

(72) 発明者 ポール・ホッジス

アメリカ合衆国95120、カリフォルニア州
サン・ホセ、モンテベルデ・ドライブ
6154

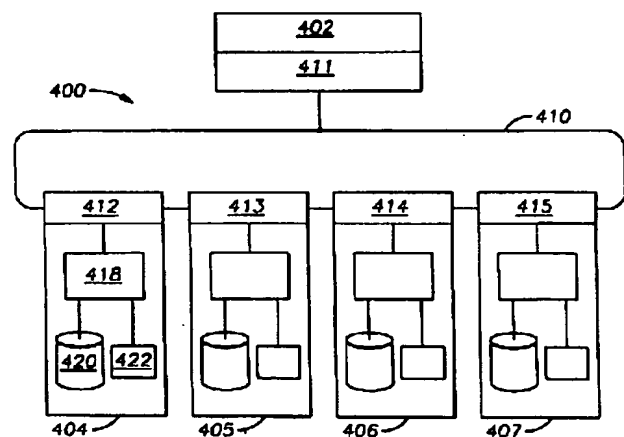
(74) 代理人 弁理士 合田 潔 (外2名)

(54) 【発明の名称】 局所XOR機能を有するデータ記憶システム

(57) 【要約】

【課題】 複数のXOR演算を記憶装置間で分散して実行するデータ記憶システムを提供する。

【解決手段】 データ記憶システムが制御装置及び複数のDASDのストリングを含み、各DASDが制御装置からのコマンドに应答して、XOR機能及び他のDASDへのデータ転送などの局所オペレーションを実行する。直列または並列バスを含むバスが、制御装置及びDASDを電氣的に相互接続する。各DASDは記憶装置、インタフェース、バッファ、及びプロセッサを含む。バッファはインタフェース及び記憶装置から選択的にデータを受信する。他のオペレーションに加え、プロセッサは制御装置からバッファ・インタフェースへのコマンドに应答して、バッファ内の選択データ項目に対してXOR演算を実行する。プロセッサは更に、演算結果を別のDASDまたは制御装置に転送しうる。



【特許請求の範囲】

【請求項1】マシン読出し可能データを記憶する装置で使用されるDASDであって、前記装置が制御装置、複数のDASD、及び前記制御装置と前記DASDを電気的に相互接続するバスを含み、前記複数のDASDのうちの1つが、前記制御装置の各オペレーションにおけるパリティDASDとして指定されるものにおいて、マシン読出し可能データを記憶する記憶装置と、前記バスに電気的に接続されるインタフェースと、前記インタフェース及び前記記憶装置に動作的に接続される少なくとも1つのバッファと、前記バッファに電気的に接続され、前記制御装置から前記インタフェースを介してコマンドを受信し、該コマンドにตอบสนองして、前記バッファ内に記憶されるデータ項目を選択し、前記選択データ項目に対してXOR演算を実行し、該XOR演算結果を前記バスを介して、前記コマンドにより指定される、前記DASDの1つまたは前記制御装置を含む宛先に送信するプロセッサと、を含む、DASD。

【請求項2】前記プロセッサがマイクロプロセッサを含む、請求項1記載のDASD。

【請求項3】前記プロセッサがASICを含む、請求項1記載のDASD。

【請求項4】前記プロセッサが個別回路要素のアセンブリから成る論理回路を含む、請求項1記載のDASD。

【請求項5】前記記憶装置が、少なくとも1つの磁気記録ディスクを含むディスク・ドライブを含む、請求項1記載のDASD。

【請求項6】前記インタフェースが直列インタフェースを含む、請求項1記載のDASD。

【請求項7】前記インタフェースが並列インタフェースを含む、請求項1記載のDASD。

【請求項8】バスにより相互接続される複数のメンバを含むデータ記憶システムにおいて使用されるDASDであって、前記メンバが前記制御装置及び複数のDASDを含むものにおいて、

前記バスに電気的に接続されるインタフェースと、データ記憶装置と、

前記インタフェース及び前記記憶装置に動作的に接続され、それらから受信されるデータを選択的に記憶するバッファと、

前記バッファに電気的に接続され、前記制御装置から前記インタフェースを介して受信されるコマンドにตอบสนองして、次のステップ、すなわち、前記バッファと前記インタフェース間でデータを選択的に転送するステップ、前記記憶装置と前記バッファ間でデータを交換するステップ、前記バッファ内に含まれる前記選択データ項目に対してXOR演算を実行するステップ、及び前記インタフェースから選択される前記メンバにデータを転送するステップの少なくとも1つを実行するようにプログラムさ

れるプロセッサと、を含む、DASD。

【請求項9】前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドにตอบสนองして、前記記憶装置から選択データを読み出し、該選択データを選択される前記メンバに転送するステップを実行するようにプログラムされる、請求項8記載のDASD。

【請求項10】前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドにตอบสนองして、前記インタフェースにより受信される前記選択データを前記記憶装置に書込むステップを実行するようにプログラムされる、請求項8記載のDASD。

【請求項11】前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドにตอบสนองして、前記記憶装置から選択データを読み出し、前記インタフェースにより受信されるデータ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送するステップを実行するようにプログラムされる、請求項8記載のDASD。

【請求項12】前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドにตอบสนองして、前記記憶装置から前記選択データを読み出し、前記インタフェースにより受信されるデータ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送し、前記インタフェースにより受信される前記データを前記記憶装置に書込むステップを実行するようにプログラムされる、請求項8記載のDASD。

【請求項13】前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドにตอบสนองして、前記記憶装置から前記選択データを読み出し、前記インタフェースにより受信されるデータ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を前記記憶装置に書込むステップを実行するようにプログラムされる、請求項8記載のDASD。

【請求項14】前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドにตอบสนองして、前記インタフェースにより受信される前記選択データを前記記憶装置に書込み、前記選択データを選択される前記メンバに転送するステップを実行するようにプログラムされる、請求項8記載のDASD。

【請求項15】前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドにตอบสนองして、前記インタフェースにより受信される2つのデータ項目に対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送し、選択される一方の前記受信データ項目を前記記憶装置に書込むステップを実行するようにプログラムされる、請求項8記載のDASD。

【請求項16】マシン読出し可能データを記憶するシステムであって、

制御装置及び複数のDASDを含む複数のメンバと、

3.

前記メンバを電氣的に相互接続してループを形成するバスと、
 を含む、前記の各DASDが、
 マシン読出し可能データを記憶する記憶装置と、
 前記バスに電氣的に接続されるインタフェースと、
 前記インタフェース及び前記記憶装置に動作的に接続され、前記インタフェース及び前記記憶装置から受信されるデータを選択的に記憶する少なくとも1つのバッファと、
 前記バッファに電氣的に接続され、前記制御装置から前記インタフェースを介してコマンドを受信し、所定のコマンドに回答して、前記バッファ内に記憶される前記データ項目を選択し、前記選択データ項目に対してXOR演算を実行し、該XOR演算結果を前記バスを介して、選択される前記メンバを含む宛先に送信するXOR発生器と、
 を含む、システム。

【請求項17】前記プロセッサがマイクロプロセッサを含む、請求項16記載のシステム。

【請求項18】前記プロセッサがASICを含む、請求項16記載のシステム。

【請求項19】前記プロセッサが個別回路要素のアセンブリから成る論理回路を含む、請求項16記載のシステム。

【請求項20】前記記憶装置が、少なくとも1つの磁気記録ディスクを含むディスク・ドライブを含む、請求項16記載のシステム。

【請求項21】前記インタフェースが直列インタフェースを含む、請求項16記載のシステム。

【請求項22】前記インタフェースが並列インタフェースを含む、請求項16記載のシステム。

【請求項23】前記複数のDASDがRAIDプロトコルに従い構成される、請求項16記載のシステム。

【請求項24】前記RAIDプロトコルがRAID-5を含む、請求項23記載のシステム。

【請求項25】複数のXOR演算の実行を要求するタスクを実行するように構成されるデータ記憶システムであって、

コマンドを転送する制御装置と、
 前記コマンドに回答して、前記XOR演算の実行をDASD間で分散することにより、前記タスクを実行するDASDアレイと、
 前記DASD及び前記制御装置を相互接続するバスと、
 を含む、データ記憶システム。

【請求項26】データ記憶システムを動作させる方法であって、
 制御装置及び複数のDASDのストリングを含む複数のメンバと、
 前記複数のメンバを電氣的に相互接続するバスと、
 を含む、各DASDが、

4.

前記バスに電氣的に接続されるインタフェースと、
 データ記憶装置と、
 前記インタフェース及び前記記憶装置に動作的に接続されるバッファと、
 前記バッファに動作的に接続されるプロセッサと、
 を含む、
 使用可能な前記DASDを識別するステップと、
 前記制御装置から前記識別されたDASDに個々にコマンドを送信するステップと、

10 前記コマンドに回答して、データが前記識別されたDASDを所定順序でデジiser・チェーン転送されるように、前記識別されたDASDを動作させるステップと、
 を含む、データがイニシエータDASDから始まり、複数の中間DASDを経由してレシーバDASDに至る前記各DASDにより、順次転送及び受信され、前記の各中間DASDが、前記デジiser・チェーンに沿って受信される前記データを前記所定順序に従う次のDASDに転送する以前に、前記受信データ項目及び別のデータ項目に対して、XOR演算を実行する、方法。

20 【請求項27】前記別のデータ項目が前記制御装置から受信される前記データ項目を含む、請求項26記載の方法。

【請求項28】前記別のデータ項目が前記中間DASDのそれぞれの前記記憶装置から読出される前記データ項目を含む、請求項26記載の方法。

【請求項29】前記制御装置からの少なくとも1つの前記所定コマンドに回答して、選択DASDの前記記憶装置から前記データを読出し、読出したデータを選択される前記メンバに転送するように、前記選択DASDをオペレートするステップを含む、請求項26記載の方法。

30 【請求項30】前記制御装置からの少なくとも1つの前記所定コマンドに回答して、前記インタフェースにより受信される前記データを前記選択DASDの前記記憶装置に書込むように、前記選択DASDをオペレートするステップを含む、請求項26記載の方法。

【請求項31】前記制御装置からの少なくとも1つの前記所定コマンドに回答して、前記選択DASDの前記記憶装置から前記選択データを読出し、前記選択DASDの前記インタフェースにより受信される前記データ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送するように、前記選択DASDをオペレートするステップを含む、請求項26記載の方法。

50 【請求項32】前記制御装置からの少なくとも1つの前記所定コマンドに回答して、前記選択DASDの前記記憶装置から前記選択データを読出し、前記選択DASDの前記インタフェースにより受信される前記データ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送し、前記選択DASDの前記インタフェースにより受信された前記デー

タを該選択DASDの前記記憶装置に書込むように、前記選択DASDをオペレートするステップを含む、請求項26記載の方法。

【請求項33】前記制御装置からの少なくとも1つの前記所定コマンドに応答して、前記選択DASDの前記記憶装置から前記選択データを読み出し、前記選択DASDの前記インタフェースにより受信される前記データ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を前記選択DASDの前記記憶装置に書込むように、前記選択DASDをオペレートするステップを含む、請求項26記載の方法。

【請求項34】前記制御装置からの少なくとも1つの前記所定コマンドに応答して、前記選択DASDの前記インタフェースにより受信される前記選択データを前記選択DASDの前記記憶装置に書込み、前記受信データを選択される前記メンバに転送するように、前記選択DASDをオペレートするステップを含む、請求項26記載の方法。

【請求項35】前記制御装置からの少なくとも1つの前記所定コマンドに応答して、前記選択DASDの前記インタフェースにより受信される2つの前記データ項目に対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送し、選択される一方の前記受信データ項目を前記選択DASDの前記記憶装置に書込むように、前記選択DASDをオペレートするステップを含む、請求項26記載の方法。

【請求項36】制御装置、DASDアレイ、及び前記DASD及び前記制御装置を相互接続するバスを含むデータ記憶システムにおいて、複数のXOR演算の実行を要求するタスクを実行するように、前記データ記憶システムを動作させる方法であって、前記制御装置から前記バスを介して前記DASDにコマンドを転送するステップと、前記コマンドに応答して、前記XOR演算の実行を前記DASD間で分散することにより、前記タスクを実行するように、前記DASDをオペレートするステップと、を含む、方法。

【請求項37】前記オペレーティング・ステップが、データを前記DASD間で選択順に順次転送するように、前記DASDをオペレートするステップを含み、前記複数のDASDが前記選択順序に従い前記データを別のDASDに転送する以前に、該データに対して前記XOR演算を実行する、請求項36記載の方法。

【請求項38】前記オペレーティング・ステップが再生オペレーションを含み、該再生オペレーションが、第1のDASDの第1の記憶ロケーションから第1のデータ・ブロックを読み出し、前記第1のデータ・ブロックを第2のDASDに転送するように、前記第1のDASDをオペレートするステップと、前記第2のDASDをオペレートするステップであつ

て、

前記第1のデータ・ブロックを前記第1のDASDから受信するステップと、

前記第2のDASDの記憶ロケーションから、前記第1のデータ・ブロックに対応する第2のデータ・ブロックを読み出すステップと、

前記第1及び前記第2のデータ・ブロックに対してXOR演算を実行し、第1のXOR結果を生成するステップと、

10 前記第1のXOR結果を第3のDASDに転送するステップと、を含む前記オペレーティング・ステップと、前記第3のDASDをオペレートするステップであつて、

前記第1のXOR結果を前記第2のDASDから受信するステップと、

前記第3のDASDの記憶ロケーションから、前記第1及び前記第2のデータ・ブロックに対応する第3のデータ・ブロックを読み出すステップと、

20 前記第1のXOR結果及び前記第3のデータ・ブロックに対してXOR演算を実行し、第2のXOR結果を生成するステップと、

前記第2のXOR結果を前記制御装置に転送するステップと、を含む前記オペレーティング・ステップと、を含む、請求項36記載の方法。

【請求項39】前記オペレーティング・ステップが復元オペレーションを含み、該復元オペレーションが、前記第1のDASDの前記第1の記憶ロケーションから第1のデータ・ブロックを読み出すように、前記第1のDASDをオペレートするステップと、

30 第2のDASDをオペレートするステップであつて、前記第1のデータ・ブロックを前記第1のDASDから受信するステップと、

前記第2のDASDの記憶ロケーションから、前記第1のデータ・ブロックに対応する前記第2のデータ・ブロックを読み出すステップと、

前記第1及び前記第2のデータ・ブロックに対してXOR演算を実行し、第1のXOR結果を生成するステップと、

40 前記第1のXOR結果を前記第3のDASDに転送するステップと、を含む前記オペレーティング・ステップと、

前記第3のDASDをオペレートするステップであつて、

前記第1のXOR結果を前記第2のDASDから受信するステップと、

前記第3のDASDの記憶ロケーションから、前記第1及び前記第2のデータ・ブロックに対応する前記第3のデータ・ブロックを読み出すステップと、

50 前記第1のXOR結果及び前記第3のデータ・ブロックに対してXOR演算を実行し、第2のXOR結果を生成

7

するステップと、
 前記第 2 の XOR 結果を第 4 の DASD に転送するステップと、を含む前記オペレーティング・ステップと、
 前記第 4 の DASD をオペレートするステップであって、
 前記第 2 の XOR 結果を前記第 3 の DASD から受信するステップと、
 前記第 2 の XOR 結果を前記第 4 の DASD の記憶ロケーションに書込むステップと、を含む前記オペレーティング・ステップと、
 を含む、請求項 36 記載の方法。
 【請求項 40】前記オペレーティング・ステップが更新オペレーションを含み、該更新オペレーションが、
 前記第 1 のデータ・ブロックを前記制御装置から前記第 1 の DASD の前記第 1 の記憶ロケーションに転送するステップと、
 前記第 1 の DASD をオペレートするステップであって、
 前記第 1 のデータ・ブロックを前記制御装置から受信するステップと、
 前記第 1 の DASD の前記記憶ロケーションから、前記第 1 のデータ・ブロックに対応する第 2 のデータ・ブロックを読出すステップと、
 前記第 1 及び前記第 2 のデータ・ブロックに対して XOR 演算を実行し、第 1 の XOR 結果を生成するステップと、
 前記第 1 の XOR 結果を前記第 2 の DASD に転送するステップと、を含む前記オペレーティング・ステップと、
 前記第 2 の DASD をオペレートするステップであって、
 前記第 1 の XOR 結果を前記第 1 の DASD から受信するステップと、
 前記第 2 の DASD の前記記憶ロケーションから、前記第 1 及び前記第 2 のデータ・ブロックに対応する前記第 3 のデータ・ブロックを読出すステップと、
 前記第 2 及び前記第 3 のデータ・ブロックに対して XOR 演算を実行し、第 2 の XOR 結果を生成するステップと、
 前記第 2 の XOR 結果を前記第 3 の DASD に転送するステップと、を含む前記オペレーティング・ステップと、
 前記第 3 の DASD をオペレートするステップであって、
 前記第 2 の XOR 結果を前記第 2 の DASD から受信するステップと、
 前記第 2 の XOR 結果を前記第 3 の DASD の記憶ロケーションに書込むステップと、を含む前記オペレーティング・ステップと、
 を含む、請求項 36 記載の方法。

8

【請求項 41】前記オペレーティング・ステップがストライプ書込みオペレーションを含み、該ストライプ書込みオペレーションが、
 第 1 のデータ・ブロックを前記制御装置から前記第 1 の DASD の前記第 1 の記憶ロケーションに転送するステップと、
 前記第 1 の DASD をオペレートするステップであって、
 前記第 1 のデータ・ブロックを前記制御装置から受信するステップと、
 前記第 1 のデータ・ブロックを前記第 1 の DASD の記憶ロケーションに書込むステップと、
 前記第 1 のデータ・ブロックを前記第 2 の DASD に転送するステップと、を含む前記オペレーティング・ステップと、
 前記第 2 の DASD をオペレートするステップであって、
 前記第 2 のデータ・ブロックを前記制御装置から受信するステップと、
 前記第 2 のデータ・ブロックを前記第 2 の DASD の記憶ロケーションに書込むステップと、
 前記第 1 及び前記第 2 のデータ・ブロックに対して XOR 演算を実行し、第 1 の XOR 結果を生成するステップと、
 前記第 1 の XOR 結果を前記第 3 の DASD に転送するステップと、を含む前記オペレーティング・ステップと、
 前記第 3 の DASD をオペレートするステップであって、
 前記第 3 のデータ・ブロックを前記制御装置から受信するステップと、
 前記第 3 のデータ・ブロックを前記第 3 の DASD の記憶ロケーションに書込むステップと、
 前記第 1 の XOR 結果を前記第 2 の DASD から受信するステップと、
 前記第 1 の XOR 結果及び前記第 3 のデータ・ブロックに対して XOR 演算を実行し、第 2 の XOR 結果を生成するステップと、
 前記第 2 の XOR 結果を前記第 4 の DASD に転送するステップと、を含む前記オペレーティング・ステップと、
 前記第 4 の DASD をオペレートするステップであって、
 前記第 2 の XOR 結果を前記第 3 の DASD から受信するステップと、
 前記第 2 の XOR 結果を前記第 4 の DASD の記憶ロケーションに書込むステップと、を含む前記オペレーティング・ステップと、
 を含む、請求項 36 記載の方法。
 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明はRAID (redundant arrays of inexpensive disks) などのデータ記憶媒体のアレイに関し、特に、制御装置と、各々が受信データを処理し、処理データを異なる記憶装置または制御装置に送信するための局所排他的論理和(XOR)発生器を含む複数のデータ記憶装置とを使用するデータ記憶システムに関する。従って、本発明のデータ記憶システムは、複数のXOR演算の実行を要求するタスクを、それらのXOR演算を記憶装置間で分散して実行するように動作する。

【0002】

【従来の技術】多くの技術者が、RAIDなどの拡張データ記憶システムを使用することにより、コンピュータの記憶能力を改善しようと努めてきた。一般的に、RAID技術は、複数の同一の直接アクセス記憶装置(DASD)を使用することにより、デジタルデータを記憶する。RAID技術の概念は既知であり、RAID装置の多くの異なる変形が開発されてきた。例えば、"A Case for Redundant Arrays of Inexpensive Disks(RAID)", Report No. UCB/CSD 87/391, December 1987 (Pattersonらによる"A Case for Redundant Arrays of Inexpensive Disks(RAID)", Proceedings of the 1988 ACM SIGMOD Conference on Management of Data, Chicago, IL, June 1988としても引用される)を参照されたい。

【0003】典型的RAID装置は、RAID制御装置と複数のDASDとを含む。各DASDはマイクロプロセッサ及び様々なメモリ・バッファなどの特定の電子回路とと共に、複数のデータ記憶ディスクを含む。各DASDの特定部分が、他のDASDに対応するパリティ・ビットを記憶するために割当てられる。しかしながら、説明を容易にするために、RAIDはしばしば、パリティ情報の記憶を専用とする特定のDASDを有するように述べられる。

【0004】RAIDシステムでは、データは通常、アレイの異なるDASDの対応ロケーションに渡り、データをストライピングすることにより記憶される。記憶データのパリティ・ビットは、データにXOR機能を適用することにより生成される。従って、あるDASDが故障すると、喪失データが、残りのDASDからのデータ及び対応パリティ・データにXOR機能を適用することにより復元される。

【0005】多くのRAID記憶システムでは、次のような多数の標準機能が存在する。

1. 順次書込み
2. 更新書込み
3. キャッシュ・データによる更新
4. 故障データDASD状況の更新
5. 故障パリティDASD状況の更新
6. 喪失データの再生(regenerate)

7. 喪失データの復元(rebuild)

8. データの読出し

【0006】順次書込みは、単に、データ及びその対応パリティ・ビットをDASDに記憶する。特に、書込みでは、DASDの対応ロケーションに渡り、データの複数ブロック・グループをストライピングし、ストライプ・データに対応するパリティ・ビットのブロックを計算し、パリティ・ブロックをパリティDASDに記憶する。パリティ・ブロックは、例えば書込みデータのブロックに対してXOR演算を実行することにより、計算される。

【0007】更新書込みは、新たなデータを"ターゲット" DASDに書込み、その新たなデータに対応するパリティ・ビットを更新する。特に更新では、新たなデータ・ブロックを制御装置からターゲットDASDに書込み、更新データを反映する更新パリティ・ブロックを計算し、更新パリティ・ブロックを記憶する。更新パリティ・ブロックは、例えば、新たな(更新)データ・ブロック、更新データ・ブロックに対応する旧データ・ブロック、及び対応する旧パリティ・ブロックにXOR演算を実行することにより計算される。

【0008】キャッシュ・データによる更新は、制御装置またはDASDの1つのキャッシュ・メモリから、ターゲットDASDへ新たなデータを書込み、データ及びそのパリティ・ビットを記憶する。より詳細には、キャッシュ・データによる更新は、キャッシュ・データ・ブロックのターゲットDASDへの書込み、更新データを反映する更新パリティ・ブロックの計算、及び更新パリティ・ブロックの記憶を含む。更新パリティ・ビットは、例えば、新たな(更新)キャッシュ・データ、旧キャッシュ・データ、及びキャッシュ・データに対応する旧パリティ・ビットに対して、XOR演算を実行することにより計算される。

【0009】故障データDASD状況の更新は、新たなデータによりあるDASDが更新されるべきであるが、そのDASDが故障している状況において発生する。データが故障DASDに書込めないのが、RAID装置がせいぜい実行できることは、本来書込まれるべきデータを反映するように、パリティDASDを更新することである。従って、このプロセスは、非故障データDASDから"旧"データ・ブロックを読出し、このデータ及び新データにもとづきパリティ・ブロックを計算し、更新パリティ・ビットをパリティDASD上に記憶する。パリティ・ビットはもちろん、旧データ及び新データにXOR演算を実行することにより計算される。

【0010】故障パリティDASD状況の更新は、データ・ブロックがターゲットDASDにおいて更新されるが、パリティDASDが故障している状況において発生する。このオペレーションでは、更新データ・ブロックが単にターゲットDASDに書込まれる。パリティDA

SDを更新するステップは省略される。

【0011】故障DASDからの喪失データの再生 (regenerating) は、故障DASDからデータを復元し、それをRAID制御装置に提供する。特に、このプロセスは全ての非故障DASDから対応データ・ブロックを読み出し、読出されたデータ・ブロックにもとづき喪失データ・ブロックを復元し、復元されたデータ・ブロックを制御装置に転送する。喪失データ・ブロックは、例えば、非故障DASDから読出されるデータ・ブロックに対して、XOR演算を実行することにより復元される。

【0012】復元 (rebuilding) は、以前に故障した (現在は使用可能な) ターゲットDASDからデータ・ブロックを復元 (reconstruct) し、復元されたデータ・ブロックをターゲットDASDに書き込み、パリティ・ブロックを更新する。

【0013】データの読出しは、複数のDASDの1つに記憶されるデータ・ブロックを読出す。

【0014】従来、技術者達はRAIDにおける複数の異なる変形を開発した。これらの多くは、ここでの議論で使用されるRAID-5機構に向けられる。RAIDシステムは通常、集中化または分散構成により構成される。集中化構成では、RAID制御装置がDASDのために多くの機能を実行し、上述の様々なRAID機能を実行する支配的な役割を演じる。この場合、制御装置はRAIDシステムの“頭脳”のごとく作用し、コマンドを個々のDASDに提供し、DASDから情報を収集し、DASD内でのデータの処理及び記憶を制御する。制御装置は例えばパリティ・ビットの生成を管理し、DASDの1つが故障した場合のデータの復元を制御する。集中化RAIDシステムの重要な特長の1つは、制御装置がドライブに書込まれるデータまたはそこから取り出されるデータに対して実行されなければならない全てのXOR機能を実行することである。

【0015】一方、分散システムは、かなりの責任を個々のDASDに帰属させる。分散システムでは、制御装置はある程度DASDを管理するが、通常、DASDの1つが、他のDASDの特定のオペレーションを管理する重要な役割を担う。本発明を理解する上で有用な複数の原理を用いる分散RAIDシステムの1つは、ここでは“外部XOR”アプローチとして参照される。すなわち、XOR機能は制御装置にとって“外部的”であるが、DASDにとっては局所的である。こうした局所計算は、喪失データの復元、パリティ計算などを支援するために、順次結合されう。図1に示されるように、外部XORアプローチは、複数のDASD104乃至107に接続される中央制御装置102を用いるRAID装置100に関わる。各DASD104乃至107は、バッファ、XOR回路、及びディスク・ドライブなどの記憶装置を含む。例えば、DASD104は、バッファ104a、XOR回路104b、及びディスク・ドライブ1

04cを含む。

【0016】図1では、各DASD104乃至107が、バス110を介して、他のDASD104乃至107及び制御装置102に接続される。外部XOR機構の下で、制御装置102はイニシエータDASDを指定し、単一のSCSIコマンドを指定イニシエータDASDに発行することにより、アレイ・オペレーションを開始する。図1の例では、イニシエータDASDはDASD107である。イニシエータDASD107は、制御装置102により要求されるオペレーションを達成するSCSIイニシエータの役割を担う。すなわち、イニシエータDASD107は、制御装置102にとって外部的にオペレーションを達成し、終了時に終了ステータスを制御装置102に報告する。一般に、エラー回復はDASDの役割である。しかしながら、DASDレベルにおいて回復不能なエラーに対しては、エラー回復責任を制御装置102に転嫁するプロシージャが存在する。

【0017】外部XORアプローチは、多数の魅力的な特長を有する。例えば、このアプローチは、RAID制御装置の作業負荷を軽減する。また、XOR機能の局所化は、特殊RAID制御装置の必要を回避しうり、特定のケースでは、RAID装置が一般のSCSI制御装置と互換にさえなりうる。

【0018】しかしながら、特定のアプリケーションでは、外部XORアプローチの下でのシステム性能が、ユーザの期待にそぐわないことがありうる。例えば、制御装置102が複数のイニシエータを指定する場合、DASD104乃至107が、同時に複数のイニシエータによる一貫性の無い制御に晒されう。特に、制御装置によるコマンドの発行と、任意のDASDによるそのコマンドの受信との間に遅延が存在しう。従って、連続的なコマンドが予測不能かつ不正な順序で受信され、実行される機会が存在しう。特定のケースでは、これはデータの保全性を低下しう。

【0019】外部XORアプローチの別の潜在的欠点は、計算効率に関わる。特にこのアプローチは、特定のRAIDオペレーションにおいて過剰なタスクを実行する。例えば、図2は順次書き込みオペレーションの間の、外部XOR装置200の振舞いを示す。外部XOR装置200は制御装置202、データDASD204乃至206、及びパリティDASD208を含む。外部XOR装置200は、順次書き込みオペレーションを次のように実行する。第1に、制御装置202が、XORデータ書き込み (XDW) コマンド204a乃至206aを、それぞれDASD204乃至206に発行する。各書き込みコマンドと共に、制御装置202は、対応するDASDに書込まれるデータ・ブロックを提供する。データ・ブロックは204b乃至206bにより示される。

【0020】コマンド及びデータ・ブロックの受信に際し、DASD204乃至206は、適切なデータをそれ

らのディスク・ドライブに書き込む。更に各DASDは、XOR機能を新たなデータ及び置換される旧データに適用する。これらのXOR演算の結果は、後にパリティ・ビットを生成するために使用されるデータを提供し、パリティ・ビットはパリティDASD208により記憶される。従って、次のステップでは、DASD204乃至206は、パリティDASD208に、XORパリティ書き込み(XPW)コマンド204d乃至206dと共に、計算されたパリティ・ビット204c乃至206cを送信する。次に、パリティDASD208は受信データから最終パリティ・ビットを計算し、これがパリティDASD208のディスク・ドライブに書き込まれる。後述の説明で示されるように、外部XORアプローチは、図2により示されるような状況において、過剰なオペレーションを要求する。

【0021】更に、外部XORアプローチの別の潜在的欠点は、特定のオペレーションが過度な負荷をイニシエータDASDに課すことである。一例として、図3に示される外部XOR装置200により実行される再生オペレーションが挙げられる。再生オペレーションは、故障DASD300上に配置されるために取り出せない“ミッシング(missing)”データ・ブロックを復元するために実行される。再生コマンドはまた再生データを制御装置202に提供する。

【0022】最初に制御装置202は、再生コマンド302をイニシエータDASDに発行する。図3の例では、イニシエータDASDはDASD208である。イニシエータDASD208は次に、ミッシング・データ・ブロックに対応するデータ・ブロックを取り出すために、読出しコマンド304乃至306を非故障DASD204乃至206に送信する。それに応答して、各DASD204乃至206は、イニシエータDASD208に要求データ・ブロック308乃至310を送信する。イニシエータDASD208はまた、自身のディスク・ドライブからも、ミッシング・データ・ブロックに対応するデータ・ブロックを取り出す。次に制御装置202がイニシエータDASD208に、XORデータ読出し(XDR)コマンドを送信する。それに応答して、イニシエータDASD208はデータ・ブロックに対してXOR演算を実行して、ミッシング・データ・ブロックを効果的に再生し、再生データ・ブロックを制御装置202に返却する。

【0023】

【発明が解決しようとする課題】上述のように、外部XORアプローチは再生オペレーションの実行において、イニシエータDASD208に大きく頼る。イニシエータDASD208はコマンドを他の各DASDに送信し、そこからデータ・ブロックを受信し、自身のデータ・ブロックの1つを取り出し、全ての抽出データ・ブロックに対してXOR演算を実行し、XOR結果を制御装

置202に返却しなければならない。結果的に、イニシエータDASD208は、外部XOR装置200のオペレーションを多大に遅延する障害を有しうる。イニシエータDASD208の重負荷により、XOR演算の実行に相当な計算時間が必要とされる。加えて、この重負荷は、イニシエータDASD208がメモリ・バッファ及びバッファ・サイクルなどの、追加のハードウェアを使用することを要求する。

【0024】更に、外部XOR構成は、制御装置からは多大な作業負荷を免除するが、イニシエータDASDに課せられる高度な責任が、事実上、イニシエータDASD自体を“制御装置”たらしめる。従って、イニシエータDASDが、他のDASDと通信し、それらを管理するためのマイクロコード及び制御回路などの、必要な知能を含まねばならない。

【0025】

【課題を解決するための手段】一般的に、本発明の1つの態様は、マシン読出し可能データを記憶するデータ記憶システムに関する。このシステムは制御装置、複数DASD、及び制御装置とDASDを電気的に相互接続するバスを含む。各DASDはマシン読出し可能データを記憶する記憶装置と、バスに電気的に接続されるインタフェースを含む。更に、各DASDはインタフェースから受信されるデータを選択的に記憶する1つ以上のバッファ、及び記憶装置を含む。各DASDは更に、制御装置からのコマンドに応答して、バッファに受信されたデータ・ブロックを選択し、選択ブロックに対してXOR演算を実行するプロセッサを含む。プロセッサはまた、その演算結果をバスを介して、別のDASDまたは制御装置などの宛先に転送しうる。

【0026】本発明のデータ記憶システムは、複数XOR演算の実行を要求するタスクを、それらXOR演算の実行をDASD間で分散して実行する。より詳細には、データ記憶システムがDASD間で、データを選択“デイジー・チェーン”順に順次的に転送するように動作する。複数DASDは、データをデイジー・チェーン内の次のDASDに転送する以前に、データに対してXOR演算を実行する。

【0027】従って、本発明はそのユーザに複数の別個の利点を提供する。第1に、本発明は以前の構成に勝る信頼性を提供する。更に、本発明は、既知の構成に勝る計算効率を提供する。本発明はまた、制御装置の作業負荷を軽減し、DASD間で処理作業負荷を分散することにより、任意のDASDの過剰作業を回避する。

【0028】

【発明の実施の形態】一般に、本発明は制御装置及び複数のDASDを含むデータ記憶装置を含み、DASDは制御装置の計算作業負荷を軽減するために、それらの間でデータを処理し、順次転送することができ、それにより計算効率及びデータ信頼性の向上を達成する。

【0029】ハードウェア要素及び相互接続：図4は本発明の好適なデータ記憶システム400を示す。システム400は好適にはRAID-5プロトコルを実現するが、RAID-1、RAID3または別のRAID構成が使用されてもよい。システム400は制御装置402を含み、これはDASD404乃至407に電気的に接続される。図示の態様では、制御装置402は、本明細書で述べられるオペレーションを実行する個別の回路要素または特定のアプリケーション向け集積回路(ASIC)を含む。代わりに、制御装置402が、本明細書で述べられる機能を達成するために好適なソフトウェアを実行するデジタル・コンピュータ・システムを含んでもよい。

【0030】好適な実施例では、DASD404乃至407及び制御装置402は、バス410により接続される。より詳細には、制御装置402はインタフェース411を含み、各DASD404乃至407はインタフェース412乃至415をそれぞれ含む。インタフェース412乃至415は、好適には直列記憶アーキテクチャなどの、SCSIコマンド・セットと互換の直列インタフェースを含む。バス410はインタフェース411乃至415を1連続ループに相互接続し、各インタフェースは2つの隣接インタフェース間に接続される。情報を2重DASDインタフェース(図示せず)間で伝達したり、バス410が故障の場合のバックアップを提供するために、2重バス(図示せず)も提供されうる。

【0031】制御装置402及びDASD404乃至407を相互接続する代わりに、他の構成も使用されうる。例えば制御装置402及びDASD404乃至407が共通バスに並列に接続されたり、ファイバ・チャネル・アービトラレーテッド・ループ・タイプの相互接続が使用されてもよい。或いは、制御装置と各DASD404乃至407間、及び任意のDASDと別のDASD間で信号を交換する特定の手段を提供する、異なる相互接続機構が使用されてもよい。

【0032】図示の態様では、各DASD404乃至407がインタフェース、プロセッサ、バッファ及び記憶装置を含む。例えば図4でDASD404はインタフェース412、プロセッサ418、データ記憶装置420及びバッファ422を含む。典型的なDASD500の構成要素が、図5に詳細に示される。DASD500はインタフェース502、プロセッサ504、データ記憶装置506、及びバッファ508を含む。インタフェース502は、第1及び第2のバッファ514、516を含み、これらはそれぞれバス410の第1の終端510及び第2の終端512に電気的に接続される。インタフェース502はまたゲート518を含み、これはバス410の終端510、512間で信号を選択的に交換し、更にプロセッサ504がバス410からの受信データにオペレーションを実行することを支援する。

【0033】プロセッサ504はバッファ514、516に電気的に接続され、それらとデータを交換する。プロセッサ504はマイクロプロセッサ、個別論理回路、またはASICを含み、データを処理し本発明に従いDASD500のオペレーションを管理する。図示の態様では、プロセッサ504は、例えばモデル80186マイクロプロセッサ、またはウェスタン・デジタル社のモデルWD33C96及びWD61C40回路の組合わせを含みうる。プロセッサ504は、記憶装置506及びバッファ508にも電気的に接続される。

【0034】図示の態様では、記憶装置506は、1つ以上の回転磁気記録媒体を有するハード・ディスク・ドライブを含みうる。或いは、DASD404乃至407が光ディスク・ドライブ、磁気テープ・ドライブ、フロッピーまたはリジッド・ディスケット、ランダム・アクセス・メモリ(RAM)、または他の好適なコンピュータ読出し可能データ記憶装置を含んでもよい。バッファ508、514及び516は、好適にはDRAMバッファを含み、バッファ508は後述される理由により、複数の従属バッファに区分化されることが望ましい。

【0035】オペレーション方法：上述のデータ記憶システムの様々な態様に加え、本発明はこうしたシステムをオペレートする方法を包含する。一般に、本発明のデータ記憶システムをオペレートするために、制御装置402(図4)はコマンド及びデータをバス410を介して、DASD404乃至407に送信する。DASD404乃至407は、データ記憶システム全体の読出し、書込み、再生、及び他のオペレーションの実行を支援するために要求される様々な局所タスクを実行して応答する。本発明の図示の態様の利点は、XOR演算がDASD404乃至407間で分散され、XOR演算を要求する所与の機能に対して、どのDASDも複数のXOR演算を実行するように要求されず、従って処理作業負荷の過度な分配を請け負わないことである。

【0036】コマンド・セット：本発明の図示の態様では、制御装置402は様々なコマンドを選択し、DASD404乃至407に発行することにより、様々な特定のオペレーション(後述)を実行する。本発明の図示の態様では、各DASDがそのゲート518により、プロセッサ504とバッファ514、516との間で信号を選択的に交換しうる。例えば、制御装置402がDASD406に対してコマンドを発行する場合、制御装置402はそのコマンドをバス410上に出力する。コマンドはインタフェース412、413及び414を介して、インタフェース415に転送される。この状況では、各インタフェース412、413、414のゲート518は、コマンドが別のDASDに向けられることを認識し、信号をバッファ514からバッファ516に転送し、事実上、バス410の終端510、512を接続する。

【0037】各コマンドは、DASD404乃至407により認識される全てのコマンドを含む“コマンド・セット”から選択される。本発明の図示の態様では、コマンド・セットが次のコマンドを含みうる。WRITE、XT、XW、XTW、XTW/NR、TW、READ、及びTRANSFER。

【0038】WRITE：このコマンドは、ターゲットDASD500に、特定の入来データをその記憶装置506に記憶するように指令する。この点に関し、“入来”データは、インタフェース502によりバス410から獲得されるデータを指す。従って、入来データは制御装置402、別のDASD、または別の制御装置（図示せず）などの発信元から発信されうる。入来データが制御装置402以外の発信元から提供される場合、制御装置402もまたTRANSFERコマンド（後述）をデータ発信元に送信する。

【0039】XT：このコマンドは、ターゲットDASD500に、特定のデータをその記憶装置506から読出し、そのデータと特定の入来データの組合わせにXORを実行し、XOR結果を2次ターゲットに転送するように指令する。ここでの説明の都合上、“2次ターゲット”は別のDASD、制御装置402、または異なる制御装置（図示せず）を含みうるものとする。WRITEコマンド同様、入来データは制御装置402、別のDASD、または別の制御装置（図示せず）などの発信元から発信されうる。入来データが制御装置402以外の発信元から提供される場合、制御装置402もまたTRANSFERコマンドをデータ発信元に送信する。

【0040】ターゲットDASD500がXTコマンドを受信すると、インタフェース502が入来データを受信し、記憶装置506がXTコマンドにより指定されるデータを読出す。これらのデータ項目は、好適には続く処理のためにバッファ508に記憶される。特に、プロセッサ504は、バッファ508内に含まれる読出されたデータ及び入来データに対して、XOR演算を実行する。次にターゲットDASD500が2次イニシエータとして、限定的役割を請け負う。特に、プロセッサ504はTRANSFERコマンド（後述）を発行して、XOR結果をバス410を介して、指定2次ターゲットに転送する。

【0041】ターゲットDASD500は、実現されるDASDの特定の態様に従い、多彩な方法により2次ターゲットを識別しうる。1つの態様では、ターゲットDASD500はDASD500に局所的に記憶されるテーブルを調査し、XOR結果の送り先を決定する。このテーブルは、ターゲットDASD500と他のDASDとの関係、制御装置402から受信されるコマンド、及びXORデータの発信元を考慮する。この情報は、ターゲットDASD500が適切な2次ターゲットを選択するために使用される。別の態様では、制御装置402か

らのコマンドが、ターゲットDASD500に情報の転送先を指令するパラメータを含む。上述の両方の態様は本開示を参考に、当業者により実現されよう。

【0042】XW：このコマンドはターゲットDASD500に、特定の入来データ及びその記憶装置506からの特定のデータに対して、XOR演算を実行し、XOR結果を記憶装置に書き込むように指令する。XTコマンド同様、入来データは制御装置402、別のDASD、または別の制御装置（図示せず）などの発信元から発信されうる。

【0043】ターゲットDASD500がXWコマンドを受信すると、インタフェース502が入来データを受信し、記憶装置506がコマンドにより指定されるデータを読出す。これらのデータ項目は、好適には続く処理のためにバッファ508に記憶される。次に、プロセッサ504が、バッファ508に含まれる2つのデータ項目に対して、XOR演算を実行し、記憶装置506がXOR結果を記憶する。

【0044】XTW：このコマンドはターゲットDASD500に、特定の入来データ及びその記憶装置506からの特定のデータに対して、XOR演算を実行し、入来データをその記憶装置506に書き込み、XOR結果を2次ターゲットに転送するように指令する。インタフェース502はバス410を介して入来データを受信し、入来データは制御装置402、別のDASD、または別の発信元（図示せず）から発信されうる。

【0045】ターゲットDASD500がXTWコマンドを受信すると、記憶装置506がコマンドにより指定されるデータを読出し、インタフェース502が入来データを受信し、プロセッサ504が2つのデータ項目に対して、XOR演算を実行する。好適には、これら2つのデータ項目はプロセッサ504を支援するために、バッファ508に記憶される。次に、記憶装置506が入来データを記憶する。次にターゲットDASD500が2次イニシエータとして、限定的役割を請け負い、プロセッサ504はTRANSFERコマンド（後述）を発行して、XOR結果をインタフェース502及びバス410を介して、指定2次ターゲットに転送する。

【0046】XTW/NR：このコマンドはターゲットDASD500に、入来データの旧ブロック及び新ブロックに対してXOR演算を実行し、新データ・ブロックをターゲットDASDの記憶装置に書き込み、XOR結果を2次ターゲットに転送するように指令する。新たな入来データ・ブロックは制御装置402から発信され、旧データ・ブロックは制御装置402または別のDASDから発信されうる。ターゲットDASDはバス410を介して、入来データの新旧両ブロックを受信する。

【0047】ターゲットDASD500がXTW/NRコマンドを受信すると、インタフェース502が2つの入来データ・ブロックを受信し、バッファ508が好適

10

20

30

40

50

にはデータ・ブロックを記憶する。次にプロセッサ 504 が 2 つのデータ・ブロックに対して XOR 演算を実行し、記憶装置 506 が新たなデータ・ブロックを記憶する。2 つのデータ・ブロックが制御装置 402 から受信される場合、新データ・ブロックは、例えばターゲット DASD 500 に転送される最初のデータ・ブロックであることから識別される。ターゲット DASD 500 が制御装置 402 から新データ・ブロックを受信し、旧データ・ブロックを別の DASD から受信する場合、ターゲット DASD の記憶装置 506 は、制御装置 402 から受信されるデータ・ブロック（すなわち新データ・ブロック）を記憶する。

【0048】記憶装置 506 が入来データから新データ・ブロックを記憶した後、ターゲット DASD 500 は 2 次イニシエータとして、限定的役割を請け負う。プロセッサ 504 は TRANSFER コマンドを発行して、XOR 結果をインタフェース 502 及びバス 410 を介して、指定 2 次ターゲットに転送する。

【0049】TW: このコマンドは DASD 500 に、入来データをその記憶装置 506 に書き込み、また入来データを 2 次ターゲットに転送するように指令する。特に、TW コマンドに回答して、インタフェース 502 が入来データを受信し、記憶装置 506 が入来データを記憶し、DASD 500 が 2 次イニシエータとして限定的役割を請け負う。プロセッサ 504 は TRANSFER コマンドを発行し、データを 2 次ターゲットに転送する。

【0050】READ: このコマンドはターゲット DASD 500 に、その記憶装置 506 の指定アドレスからデータを読み出し、読み出したデータをバス 410 を介して 2 次ターゲットに転送するように指令する。READ コマンドは、データを読み出す記憶装置 506 のアドレスを指定する。記憶装置 506 から読み出されたデータは、図示の態様では、バッファ 508 に転送され、次にインタフェース 502 を介して、2 次ターゲットに転送される。

【0051】TRANSFER: このコマンドはデータ発信元により発行され、ターゲット DASD 500 に、データ発信元からのデータをその記憶装置 506 に受信するように準備させる。2 次ターゲットが転送データを受信すると、2 次ターゲットは、以前に制御装置 402 により指令されたオペレーションを完了することができる。より詳細には、TRANSFER コマンドは第 1 の DASD により、制御装置 402 により以前に発行されたコマンドを完了するために使用され、それにより、第 2 の DASD が第 1 の DASD からデータを受信することを支援する。特に、この態様では、第 1 の DASD が TRANSFER コマンドを第 2 の DASD に送信し、データが第 1 の DASD から第 2 の DASD へ転送されることを事前に警告する。

【0052】包括的 (comprehensive) マスタ・オペレーション: 上述のコマンド・セットは、特定の結果を獲得するために、個々の DASD に転送される個々のコマンドのグループを含む。より重要な点は、これらのコマンドの内の選択コマンドを共通体系として複数 DASD に発行することにより、制御装置 402 が DASD 内で包括的マスタ・オペレーションを開始することができ、実際に最小の作業負荷が制御装置 402 により実行されることである。本発明の重要な特長の 1 つは、以降で詳述されるように、特定のマスタ・オペレーションが DASD 404 乃至 407 を通じて、データの“デジー・チェーン”転送を有利に使用しうることである。デジー・チェーン・データ転送により、データは各 DASD を通じて直列に渡される。すなわち、イニシエータ DASD から開始し、複数の中間 DASD を通過し、最終的に受信側 DASD に到来する。好適には、データはデジー・チェーン内の各 DASD を通じて 1 度だけ転送される。後述されるように、特定のケースでは、DASD がデータをその転送以前に処理し、また他のケースでは、DASD がデータを処理すること無く、単に転送する。すなわち、コンジット (conduit) として機能する。この処理の例として、各中間 DASD が、デジー・チェーン内の直前の DASD から受信されるデータ、並びに制御装置 402 またはその中間 DASD の記憶装置から受信される別のデータ項目に対して、XOR 演算を実行する。このように、本発明のデジー・チェーン転送は DASD 間の XOR タスクについて述べ、それにより制御装置 402 の負荷を軽減し、どの DASD も不均等な負荷を負わないことを保証する。

【0053】図 6 乃至図 11 に示されるように、これらの包括的マスタ・オペレーションには、UPDATE、REGENRATE、REBUILD、UPDATE (CACHED DATA)、UPDATE (FAILED DASD)、及び WRITE STRIPE が含まれる。

【0054】UPDATE (更新): このオペレーションは一般に、新たなデータ・ブロックをターゲット DASD に書き込み、新たなデータに対応するパリティ・ビットを更新する。更新書き込み (UPDATE WRITE) オペレーションについて、図 6 を参照して詳細に説明しよう。図 6 は、制御装置 601 及び DASD 602 乃至 605 を含むデータ記憶システム 600 を示す。DASD 605 はパリティ DASD として指定される。しかしながら、各 DASD の特定部分を、他のディスク DASD に対応するパリティ・ビットを記憶するように割当ててもよい。ここでは説明の都合上、システム 600 が、パリティ情報を専用に記憶する特定の DASD 605 を有するものとする。

【0055】制御装置 601 は、XTW コマンド 608 及び新たなデータ・ブロック 610 を DASD 602 に、また XW コマンド 609 を DASD 605 に、次のシーケンスの実行に適切なパラメータと一緒に送信する

ことにより、更新オペレーションを開始する。上述のXTWコマンドに従い、DASD602が新たなデータ610及びその記憶装置からの対応データに対して、XOR演算を実行し、XOR結果615を生成する。次にDASD602は新たなデータ610をその記憶装置に書き込み、XOR結果をパリティDASD605に転送する。パリティDASD605は受信XWコマンド609に従い、受信XOR結果615及びその記憶装置からの読出しデータに対してXOR演算を実行し、このXOR演算結果をその記憶装置に書き込む。

【0056】このように、新たなデータ・ブロック610がDASD602に書き込まれ、対応パリティがパリティDASD605内で更新される。

【0057】REGENERATE（再生）：REGENERATEコマンドが図7に示され、ここでデータ記憶システム700は制御装置701及びDASD702乃至706を含み、DASD706は故障している。制御装置701は、READコマンド708をDASD702に、またXTコマンド710乃至712をDASD703乃至705に、次のシーケンスの実行に適切なパラメータと一緒に送信することにより、再生オペレーションを開始する。上述のREADコマンドに従い、DASD702がコマンド708により要求されるデータ・ブロックをその記憶装置から読出し、それをDASD703に転送する。各DASD703乃至704が別のDASDからデータ・ブロックを受信すると、それらはそのデータ・ブロック及びその記憶装置からの対応データに対してXOR演算を実行し、XOR結果を次のDASDに順番に転送する。DASD705の場合には、XOR結果は制御装置701に返送される。コマンド708、710乃至712に付随して制御装置701により発行されるパラメータは、DASD702乃至705間のデータ転送の順序を決定する（例えばDASD702からDASD703へ、次にDASD704へなど）。

【0058】このように、故障DASD706の所望のロケーションからのデータが、DASD702乃至705内のデジター・チェーンXOR計算により復元され、制御装置701に提供される。

【0059】REBUILD（復元）：REBUILDコマンドが図8に示され、ここでデータ記憶システム800は制御装置801及びDASD802乃至805を含む。DASD805は、復元されなければならないデータを有するドライブとして指定される。制御装置801は、READコマンド808をDASD802に送信し、XTコマンド810乃至811をDASD803乃至804に送信し、WRITEコマンド814をDASD805に送信することにより、復元オペレーションを開始する。全てのコマンドは、続くシーケンスの実行に適切なパラメータを添付される。上述のREADコマンドに従い、DASD802はREADコマンド808により要求される

データを、その記憶装置から読出し、それをDASD803に転送する。DASD803がデータ・ブロックをDASD802から受信すると、これはこのデータ・ブロック及びその記憶装置からの対応データにXOR演算を実行し、XOR結果を次のDASD804に順番に転送する。DASD804はDASD803と同様なオペレーションを実行する。DASD805は単にXOR結果をDASD804から受信し、その結果を自身の記憶装置（すなわちDASD805の記憶装置）に記憶する。

【0060】このように、DASD805の所望のロケーションからのデータが、DASD803乃至804のデジター・チェーンXOR計算により復元され、DASD805の適切なロケーションに書き戻される。上述の再生オペレーションと同様、コマンド808、810乃至811及び814に付随して、制御装置801により発行されるパラメータが、DASD802乃至805間のデータ転送の順序を決定する。

【0061】UPDATE(CACHED DATA)（キャッシュ・データによる更新）：UPDATE(CACHED DATA)コマンドが図9に示され、ここでデータ記憶システム900は、制御装置901及びDASD902乃至905を含む。DASD905はパリティDASDとして指定される。しかしながら、各DASDの特定部分を、他のDASDに対応するパリティ・ビットを記憶するように割当ててもよい。ここでは説明の都合上、システム900が、パリティ情報を専用に記憶する特定のDASD905を有するものとする。

【0062】キャッシュ・データによる更新オペレーションは、一般に、新たなデータ・ブロックをターゲットDASD902に記憶することにより、ターゲットDASDを更新し、それに従いパリティDASD905を更新する。しかしながら、図6の更新オペレーションと異なり、新たなデータ・ブロックにより置換される旧データ・ブロックが、以前に制御装置901にキャッシュされている。従って、ターゲットDASD902は旧データをその記憶装置から読出す必要はなく、オペレーションはより迅速に実行される。

【0063】制御装置901は、XTW/NRコマンド908、新たなデータ・ブロック912、及び旧データ・ブロック913をDASD902に送信することにより、キャッシュ・データによる更新オペレーションを開始する。制御装置901はXWコマンド910をパリティDASD905に送信する。これらのコマンドは、続くシーケンスの実行に適切なパラメータと一緒に送信される。上述のXTW/NRコマンドに従い、DASD902がデータ・ブロック912乃至913に対してXOR演算を実行し、新たなデータ・ブロック912をDASD902の記憶装置に書き込み、XOR結果をパリティDASD905に転送する。パリティDASD905は

10

20

30

40

50

XWコマンド(上述)910に従い、受信XOR結果及びその記憶装置からの対応データに対してXOR演算を実行し、最終XOR結果を計算する。次にパリティDASD905が最終XOR結果を、その記憶装置の対応ロケーションに記憶する。

【0064】このように、新たなデータ・ブロック912がDASD902に書込まれ、対応パリティがパリティDASD905内で更新される。

【0065】UPDATE(FAILED DASD)(故障DASD状況の更新)：このオペレーションは一般に、DASDの故障のために更新できないデータ・ブロックに対応するパリティ・ビットを更新する。図10を参照して、故障DASD状況の更新オペレーションをより詳しく説明しよう。図10において、データ記憶システム1000は、制御装置1001及びDASD1002乃至1005を含む。DASD1005はパリティDASDとして指定される。しかしながら、各DASDの特定部分を、他のDASDに対応するパリティ・ビットを記憶するように割当ててもよい。ここでは説明の都合上、システム1000が、パリティ情報を専用に記憶する特定のDASD1005を有するものとする。

【0066】制御装置1001は、XTコマンド1008及び新データ・ブロック1009をターゲットDASD1002に送信し、XTコマンド1010乃至1011をDASD1003乃至1004に送信し、WRITEコマンド1012をDASD1005に送信することにより、故障DASD状況の更新オペレーションを開始する。全てのコマンドは、続くシーケンスの実行に適切なパラメータを付随される。上述のXTコマンドに従い、最初のDASD1002は、新たなデータ1009及びその記憶装置からの対応データに対してXOR演算を実行し、XOR結果をDASD1003に転送する。

【0067】同様に、DASD1003乃至1004は、それぞれDASD1002、1003から受信されるデータ、及びそれらの記憶装置からの対応データに対して、XOR演算を実行する。これらのXOR結果は、図10に示されるように、それぞれDASD1004及び1005に転送される。上述のWRITEコマンドに従い、パリティDASD1005がDASD1004からXOR結果を受信し、それをDASD1005の記憶装置に書込むことにより、この結果を記憶する。

【0068】このように、新たなデータ・ブロック1009が故障DASD1006に書込まれないものの、DASD1002乃至1004のデジター・チェインXOR計算により、パリティDASD1005が更新される。

【0069】WRITE STRIPE(ストライプ書込み)：このオペレーションは一般に、新データ・ブロックの“ストライプ”をDASDに書込み、新データ・ブロックに対応するパリティ・ビットを更新する。図11を参照し

て、ストライプ書込みオペレーションを詳細に説明しよう。データ記憶システム1100は、制御装置1101及びDASD1102乃至1105を含み、DASD1105はパリティDASDとして指定される。しかしながら、各DASDの特定部分を、他のDASDに対応するパリティ・ビットを記憶するように割当ててもよい。ここでは説明の都合上、システム1100が、パリティ情報を専用に記憶する特定のDASD1105を有するものとする。

【0070】制御装置1101は、TWコマンド1108及び新データ・ブロック1109をDASD1102に送信し、XTW/NRコマンド1110及び新データ・ブロック1111をDASD1103に送信し、XTW/NRコマンド1112及び新データ・ブロック1113をDASD1104に送信し、WRITEコマンド1114をDASD1105に送信することにより、ストライプ書込みオペレーションを開始する。これらのコマンドは、続くシーケンスの実行に適切なパラメータと一緒に送信される。上述のTWコマンドに従い、DASD1102はデータ・ブロック1109を受信し、それをDASD1102に対応する記憶装置に記憶する。DASD1102は更に受信データ・ブロック1109をDASD1103に転送する。上述のXTW/NRコマンドに従い、DASD1103は、新データ・ブロック1111及びDASD1102から転送されるデータ・ブロックを受信し、これら2つのデータ・ブロックに対してXOR演算を実行する。DASD1103もまた新データ・ブロック1111をその記憶装置に書込み、そのXOR結果をDASD1104に転送する。同様に、DASD1104は、新データ・ブロック1113及びDASD1103から転送されるデータ・ブロックを受信し、これら2つのデータ・ブロックに対してXOR演算を実行する。DASD1104もまた新データ・ブロック1113をその記憶装置に書込み、そのXOR結果をDASD1105に転送する。WRITEコマンド1114に従い、パリティDASD1105がDASD1104から受信されるXOR結果を、パリティDASD1105の記憶装置に書込む。

【0071】このように、各データ・ブロックが対応DASDドライブ1102乃至1104に書込まれ、対応パリティ・ビットがDASD1102乃至1104のデジター・チェインXOR計算により更新され、更新パリティ・ビットがパリティDASD1105に記憶される。

【0072】他の包括的マスタオペレーション：上述の包括的マスタ・オペレーションに加え、ここで述べられるコマンド・セットが、本発明のデータ記憶システムと一緒に、様々な他のオペレーションを実行するために使用される。更に、新たなコマンド及びここで述べられるコマンドの適応が、本発明の範囲内で、データ記憶シ

システムにおいて実現されうる。こうしたオペレーション及びコマンドは、当業者の能力の範囲内で達成されよう。また、ここでは本発明の特定の実施例についてのみ述べてきたが、当業者には本発明の範囲から逸脱すること無しに、様々な変更及び他の実施例が実現可能であることが理解されよう。

【0073】まとめとして、本発明の構成に関して以下の事項を開示する。

【0074】(1) マシン読出し可能データを記憶する装置で使用されるDASDであって、前記装置が制御装置、複数のDASD、及び前記制御装置と前記DASDを電気的に相互接続するバスを含み、前記複数のDASDのうちの1つが、前記制御装置の各オペレーションにおけるパリティDASDとして指定されるものにおいて、マシン読出し可能データを記憶する記憶装置と、前記バスに電気的に接続されるインタフェースと、前記インタフェース及び前記記憶装置に動作的に接続される少なくとも1つのバッファと、前記バッファに電気的に接続され、前記制御装置から前記インタフェースを介してコマンドを受信し、該コマンドに应答して、前記バッファ内に記憶されるデータ項目を選択し、前記選択データ項目に対してXOR演算を実行し、該XOR演算結果を前記バスを介して、前記コマンドにより指定される、前記DASDの1つまたは前記制御装置を含む宛先に送信するプロセッサと、を含む、DASD。

(2) 前記プロセッサがマイクロプロセッサを含む、上記(1)記載のDASD。

(3) 前記プロセッサがASICを含む、上記(1)記載のDASD。

(4) 前記プロセッサが個別回路要素のアセンブリから成る論理回路を含む、上記(1)記載のDASD。

(5) 前記記憶装置が、少なくとも1つの磁気記録ディスクを含むディスク・ドライブを含む、上記(1)記載のDASD。

(6) 前記インタフェースが直列インタフェースを含む、上記(1)記載のDASD。

(7) 前記インタフェースが並列インタフェースを含む、上記(1)記載のDASD。

(8) バスにより相互接続される複数のメンバを含むデータ記憶システムにおいて使用されるDASDであって、前記メンバが前記制御装置及び前記複数のDASDを含むものにおいて、前記バスに電気的に接続されるインタフェースと、データ記憶装置と、前記インタフェース及び前記記憶装置に動作的に接続され、それらから受信されるデータを選択的に記憶するバッファと、前記バッファに電気的に接続され、前記制御装置から前記インタフェースを介して受信されるコマンドに应答して、次のステップ、すなわち、前記バッファと前記インタフェース間でデータを選択的に転送するステップ、前記記憶装置と前記バッファ間でデータを交換するステップ、前

記バッファ内に含まれる前記選択データ項目に対してXOR演算を実行するステップ、及び前記インタフェースから選択される前記メンバにデータを転送するステップの少なくとも1つを実行するようにプログラムされるプロセッサと、を含む、DASD。

(9) 前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドに应答して、前記記憶装置から選択データを読み出し、該選択データを選択される前記メンバに転送するステップを実行するようにプログラムされる、上記(8)記載のDASD。

(10) 前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドに应答して、前記インタフェースにより受信される前記選択データを前記記憶装置に書込むステップを実行するようにプログラムされる、上記(8)記載のDASD。

(11) 前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドに应答して、前記記憶装置から選択データを読み出し、前記インタフェースにより受信されるデータ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送するステップを実行するようにプログラムされる、上記(8)記載のDASD。

(12) 前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドに应答して、前記記憶装置から前記選択データを読み出し、前記インタフェースにより受信されるデータ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送し、前記インタフェースにより受信される前記データを前記記憶装置に書込むステップを実行するようにプログラムされる、上記(8)記載のDASD。

(13) 前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドに应答して、前記記憶装置から前記選択データを読み出し、前記インタフェースにより受信されるデータ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を前記記憶装置に書込むステップを実行するようにプログラムされる、上記

(8)記載のDASD。

(14) 前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドに应答して、前記インタフェースにより受信される前記選択データを前記記憶装置に書込み、前記選択データを選択される前記メンバに転送するステップを実行するようにプログラムされる、上記(8)記載のDASD。

(15) 前記プロセッサが、前記制御装置からの少なくとも1つの所定のコマンドに应答して、前記インタフェースにより受信される2つのデータ項目に対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送し、選択される一方の前記受信データ項目を前記記憶装置に書込むステップを実行するようにプログラムされる、上記(8)記載のDASD。

(16) マシン読出し可能データを記憶するシステムであって、制御装置及び複数のDASDを含む複数のメンバと、前記メンバを電氣的に相互接続してループを形成するバスと、を含み、前記の各DASDが、マシン読出し可能データを記憶する記憶装置と、前記バスに電氣的に接続されるインタフェースと、前記インタフェース及び前記記憶装置に動作的に接続され、前記インタフェース及び前記記憶装置から受信されるデータを選択的に記憶する少なくとも1つのバッファと、前記バッファに電氣的に接続され、前記制御装置から前記インタフェースを介してコマンドを受信し、所定のコマンドにตอบสนองして、前記バッファ内に記憶される前記データ項目を選択し、前記選択データ項目に対してXOR演算を実行し、該XOR演算結果を前記バスを介して、選択される前記メンバを含む宛先に送信するXOR発生器と、を含む、システム。

(17) 前記プロセッサがマイクロプロセッサを含む、上記(16)記載のシステム。

(18) 前記プロセッサがASICを含む、上記(16)記載のシステム。

(19) 前記プロセッサが個別回路要素のアセンブリから成る論理回路を含む、上記(16)記載のシステム。

(20) 前記記憶装置が、少なくとも1つの磁気記録ディスクを含むディスク・ドライブを含む、上記(16)記載のシステム。

(21) 前記インタフェースが直列インタフェースを含む、上記(16)記載のシステム。

(22) 前記インタフェースが並列インタフェースを含む、上記(16)記載のシステム。

(23) 前記複数のDASDがRAIDプロトコルに従い構成される、上記(16)記載のシステム。

(24) 前記RAIDプロトコルがRAID-5を含む、上記(23)記載のシステム。

(25) 複数のXOR演算の実行を要求するタスクを実行するように構成されるデータ記憶システムであって、コマンドを転送する制御装置と、前記コマンドにตอบสนองして、前記XOR演算の実行をDASD間で分散することにより、前記タスクを実行するDASDアレイと、前記DASD及び前記制御装置を相互接続するバスと、を含む、データ記憶システム。

(26) データ記憶システムを動作させる方法であって、制御装置及び複数のDASDのストリングを含む複数のメンバと、前記複数のメンバを電氣的に相互接続するバスと、を含み、各DASDが、前記バスに電氣的に接続されるインタフェースと、データ記憶装置と、前記インタフェース及び前記記憶装置に動作的に接続されるバッファと、前記バッファに動作的に接続されるプロセッサと、を含み、使用可能な前記DASDを識別するステップと、前記制御装置から前記識別されたDASDに個々にコマンドを送信するステップと、前記コマンドに

ตอบสนองして、データが前記識別されたDASDを所定順序でデジiser・チェーン転送されるように、前記識別されたDASDを動作させるステップと、を含み、データがイニシエータDASDから始まり、複数の中間DASDを経由してレシーバDASDに至る前記各DASDにより、順次転送及び受信され、前記の各中間DASDが、前記デジiser・チェーンに沿って受信される前記データを前記所定順序に従う次のDASDに転送する以前に、前記受信データ項目及び別のデータ項目に対して、XOR演算を実行する、方法。

(27) 前記別のデータ項目が前記制御装置から受信される前記データ項目を含む、上記(26)記載の方法。

(28) 前記別のデータ項目が前記中間DASDのそれぞれの前記記憶装置から読出される前記データ項目を含む、上記(26)記載の方法。

(29) 前記制御装置からの少なくとも1つの前記所定コマンドにตอบสนองして、選択DASDの前記記憶装置から前記データを読み出し、読み出したデータを選択される前記メンバに転送するように、前記選択DASDをオペレートするステップを含む、上記(26)記載の方法。

(30) 前記制御装置からの少なくとも1つの前記所定コマンドにตอบสนองして、前記インタフェースにより受信される前記データを前記選択DASDの前記記憶装置に書込むように、前記選択DASDをオペレートするステップを含む、上記(26)記載の方法。

(31) 前記制御装置からの少なくとも1つの前記所定コマンドにตอบสนองして、前記選択DASDの前記記憶装置から前記選択データを読み出し、前記選択DASDの前記インタフェースにより受信される前記データ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送するように、前記選択DASDをオペレートするステップを含む、上記(26)記載の方法。

(32) 前記制御装置からの少なくとも1つの前記所定コマンドにตอบสนองして、前記選択DASDの前記記憶装置から前記選択データを読み出し、前記選択DASDの前記インタフェースにより受信される前記データ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送し、前記選択DASDの前記インタフェースにより受信された前記データを該選択DASDの前記記憶装置に書込むように、前記選択DASDをオペレートするステップを含む、上記(26)記載の方法。

(33) 前記制御装置からの少なくとも1つの前記所定コマンドにตอบสนองして、前記選択DASDの前記記憶装置から前記選択データを読み出し、前記選択DASDの前記インタフェースにより受信される前記データ及び前記選択データに対してXOR演算を実行し、該XOR演算結果を前記選択DASDの前記記憶装置に書込むように、前記選択DASDをオペレートするステップを含む、上

記(26)記載の方法。

(34) 前記制御装置からの少なくとも1つの前記所定コマンドに回答して、前記選択DASDの前記インタフェースにより受信される前記選択データを前記選択DASDの前記記憶装置に書き込み、前記受信データを選択される前記メンバに転送するように、前記選択DASDをオペレートするステップを含む、上記(26)記載の方法。

(35) 前記制御装置からの少なくとも1つの前記所定コマンドに回答して、前記選択DASDの前記インタフェースにより受信される2つの前記データ項目に対してXOR演算を実行し、該XOR演算結果を選択される前記メンバに転送し、選択される一方の前記受信データ項目を前記選択DASDの前記記憶装置に書き込むように、前記選択DASDをオペレートするステップを含む、上記(26)記載の方法。

(36) 制御装置、DASDアレイ、及び前記DASD及び前記制御装置を相互接続するバスを含むデータ記憶システムにおいて、複数のXOR演算の実行を要求するタスクを実行するように、前記データ記憶システムを動作させる方法であって、前記制御装置から前記バスを介して前記DASDにコマンドを転送するステップと、前記コマンドに回答して、前記XOR演算の実行を前記DASD間で分散することにより、前記タスクを実行するように、前記DASDをオペレートするステップと、を含む、方法。

(37) 前記オペレーティング・ステップが、データを前記DASD間で選択順に順次転送するように、前記DASDをオペレートするステップを含み、前記複数のDASDが前記選択順序に従い前記データを別のDASDに転送する以前に、該データに対して前記XOR演算を実行する、上記(36)記載の方法。

(38) 前記オペレーティング・ステップが再生オペレーションを含み、該再生オペレーションが、第1のDASDの第1の記憶ロケーションから第1のデータ・ブロックを読み出し、前記第1のデータ・ブロックを第2のDASDに転送するように、前記第1のDASDをオペレートするステップと、前記第2のDASDをオペレートするステップであって、前記第1のデータ・ブロックを前記第1のDASDから受信するステップと、前記第2のDASDの記憶ロケーションから、前記第1のデータ・ブロックに対応する第2のデータ・ブロックを読み出すステップと、前記第1及び前記第2のデータ・ブロックに対してXOR演算を実行し、第1のXOR結果を生成するステップと、前記第1のXOR結果を第3のDASDに転送するステップと、を含む前記オペレーティング・ステップと、前記第3のDASDをオペレートするステップであって、前記第1のXOR結果を前記第2のDASDから受信するステップと、前記第3のDASDの記憶ロケーションから、前記第1及び前記第2のデータ

・ブロックに対応する第3のデータ・ブロックを読み出すステップと、前記第1のXOR結果及び前記第3のデータ・ブロックに対してXOR演算を実行し、第2のXOR結果を生成するステップと、前記第2のXOR結果を前記制御装置に転送するステップと、を含む前記オペレーティング・ステップと、を含む、上記(36)記載の方法。

(39) 前記オペレーティング・ステップが復元オペレーションを含み、該復元オペレーションが、前記第1のDASDの前記第1の記憶ロケーションから第1のデータ・ブロックを読み出すように、前記第1のDASDをオペレートするステップと、第2のDASDをオペレートするステップであって、前記第1のデータ・ブロックを前記第1のDASDから受信するステップと、前記第2のDASDの記憶ロケーションから、前記第1のデータ・ブロックに対応する前記第2のデータ・ブロックを読み出すステップと、前記第1及び前記第2のデータ・ブロックに対してXOR演算を実行し、第1のXOR結果を生成するステップと、前記第1のXOR結果を前記第3のDASDに転送するステップと、を含む前記オペレーティング・ステップと、前記第3のDASDをオペレートするステップであって、前記第1のXOR結果を前記第2のDASDから受信するステップと、前記第3のDASDの記憶ロケーションから、前記第1及び前記第2のデータ・ブロックに対応する前記第3のデータ・ブロックを読み出すステップと、前記第1のXOR結果及び前記第3のデータ・ブロックに対してXOR演算を実行し、第2のXOR結果を生成するステップと、前記第2のXOR結果を第4のDASDに転送するステップと、を含む前記オペレーティング・ステップと、前記第4のDASDをオペレートするステップであって、前記第2のXOR結果を前記第3のDASDから受信するステップと、前記第2のXOR結果を前記第4のDASDの記憶ロケーションに書き込むステップと、を含む前記オペレーティング・ステップと、を含む、上記(36)記載の方法。

(40) 前記オペレーティング・ステップが更新オペレーションを含み、該更新オペレーションが、前記第1のデータ・ブロックを前記制御装置から前記第1のDASDの前記第1の記憶ロケーションに転送するステップと、前記第1のDASDをオペレートするステップであって、前記第1のデータ・ブロックを前記制御装置から受信するステップと、前記第1のDASDの前記記憶ロケーションから、前記第1のデータ・ブロックに対応する第2のデータ・ブロックを読み出すステップと、前記第1及び前記第2のデータ・ブロックに対してXOR演算を実行し、第1のXOR結果を生成するステップと、前記第1のXOR結果を前記第2のDASDに転送するステップと、を含む前記オペレーティング・ステップと、前記第2のDASDをオペレートするステップであって、

て、前記第1のXOR結果を前記第1のDASDから受信するステップと、前記第2のDASDの前記記憶ロケーションから、前記第1及び前記第2のデータ・ブロックに対応する前記第3のデータ・ブロックを読出すステップと、前記第2及び前記第3のデータ・ブロックに対してXOR演算を実行し、第2のXOR結果を生成するステップと、前記第2のXOR結果を前記第3のDASDに転送するステップと、を含む前記オペレーティング・ステップと、前記第3のDASDをオペレートするステップであって、前記第2のXOR結果を前記第2のDASDから受信するステップと、前記第2のXOR結果を前記第3のDASDの記憶ロケーションに書き込むステップと、を含む前記オペレーティング・ステップと、を含む、上記(36)記載の方法。

(41) 前記オペレーティング・ステップがストライプ書き込みオペレーションを含み、該ストライプ書き込みオペレーションが、第1のデータ・ブロックを前記制御装置から前記第1のDASDの前記第1の記憶ロケーションに転送するステップと、前記第1のDASDをオペレートするステップであって、前記第1のデータ・ブロックを前記制御装置から受信するステップと、前記第1のデータ・ブロックを前記第1のDASDの記憶ロケーションに書き込むステップと、前記第1のデータ・ブロックを前記第2のDASDに転送するステップと、を含む前記オペレーティング・ステップと、前記第2のDASDをオペレートするステップであって、前記第2のデータ・ブロックを前記制御装置から受信するステップと、前記第2のデータ・ブロックを前記第2のDASDの記憶ロケーションに書き込むステップと、前記第1及び前記第2のデータ・ブロックに対してXOR演算を実行し、第1のXOR結果を生成するステップと、前記第1のXOR結果を前記第3のDASDに転送するステップと、を含む前記オペレーティング・ステップと、前記第3のDASDをオペレートするステップであって、前記第3のデータ・ブロックを前記制御装置から受信するステップと、前記第3のデータ・ブロックを前記第3のDASDの記憶ロケーションに書き込むステップと、前記第1のXOR結果を前記第2のDASDから受信するステップと、前記第1のXOR結果及び前記第3のデータ・ブロックに対してXOR演算を実行し、第2のXOR結果を生成するステップと、前記第2のXOR結果を前記第4のDASDに転送するステップと、を含む前記オペレーティング・ステップと、前記第4のDASDをオペレートするステップであって、前記第2のXOR結果を前記第3のDASDから受信するステップと、前記第2のXOR結果を前記第4のDASDの記憶ロケーションに書き込むステップと、を含む前記オペレーティング・ステップと、を含む、上記(36)記載の方法。

【図面の簡単な説明】

【図1】“外部XOR”アーキテクチャによるRAID装

置100のブロック図である。

【図2】“順次書き込み”オペレーションを実行するために要求されるオペレーションを示す、外部XOR RAID装置200のデータ・フロー図である。

【図3】“再生”オペレーションを実行するために要求されるオペレーションを示す、外部XOR RAID装置200のデータ・フロー図である。

【図4】本発明によるデータ記憶システムのブロック図である。

【図5】本発明によるDASDのハードウェア要素と相互接続を示す詳細ブロック図である。

【図6】本発明による更新タスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【図7】本発明による再生タスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【図8】本発明による復元タスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【図9】本発明によるキャッシュ・データによる更新タスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【図10】本発明による故障DASD状況の更新タスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【図11】本発明によるストライプ書き込みタスクを実行するためのオペレーション・シーケンスを示すデータ・フロー図である。

【符号の説明】

100 RAID装置

102 中央制御装置

104、105、106、204、205、206、4

04、405、406、407、500、602、60

3、604、702、703、704、705、80

2、803、804、805、902、903、90

4、1002、1003、1004、1102、110

3、1104 DASD

107、208、605、905、1005、1105

40 パリティDASD

110、410 バス

200 外部XOR装置

202、402、601、701、801、901、1

001、1101 制御装置

300、706、1006 故障DASD

308、309、310、610、912、913、1

009、1109、1111、1113 データ・ブ

ロック

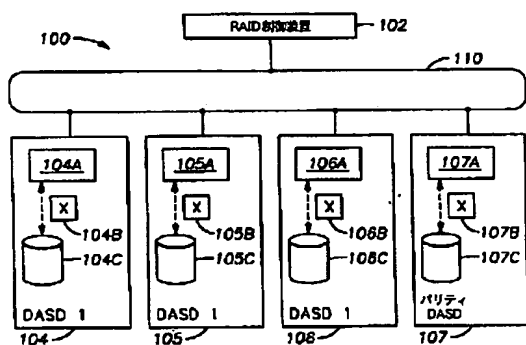
400、600、700、800、900、1000、

50 1100 データ記憶システム

33

411、412、413、414、415、502 インタフェース
 418、504 プロセッサ
 420、506 データ記憶装置

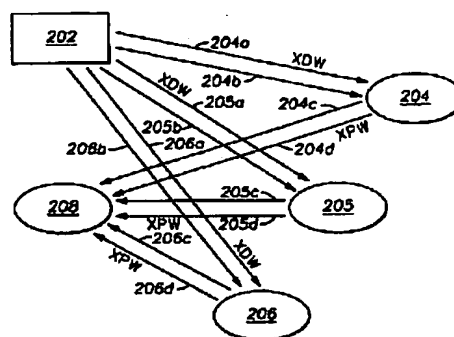
【図 1】



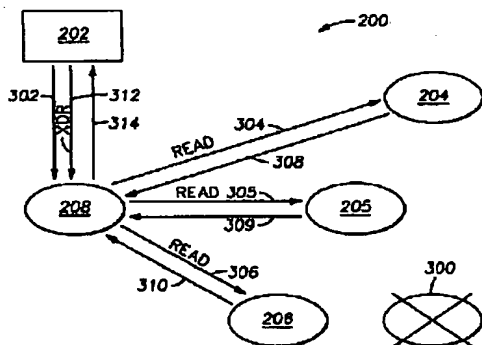
34

422、508、514、516 バッファ
 510、512 終端
 518 ゲート

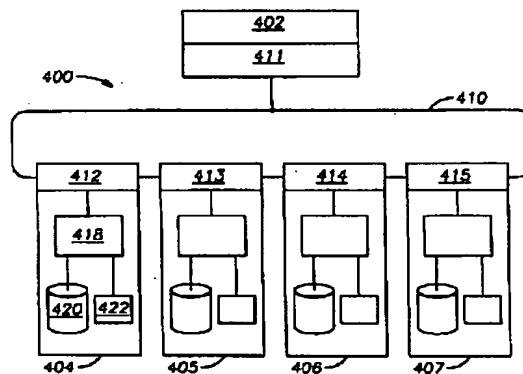
【図 2】



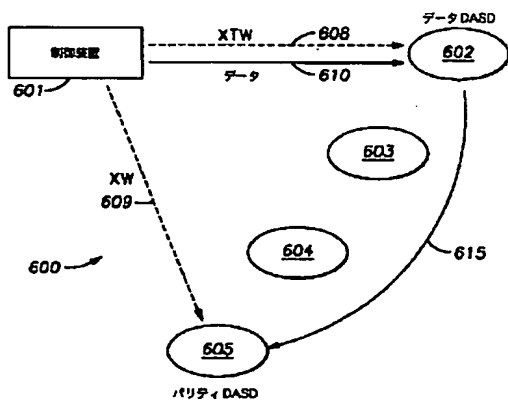
【図 3】



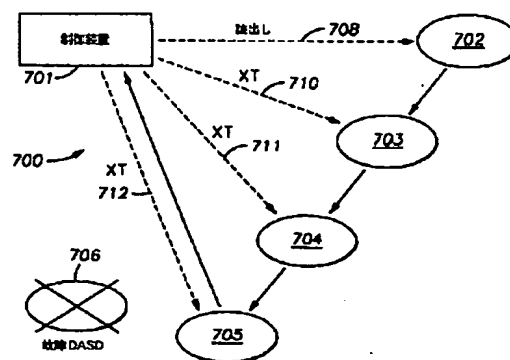
【図 4】



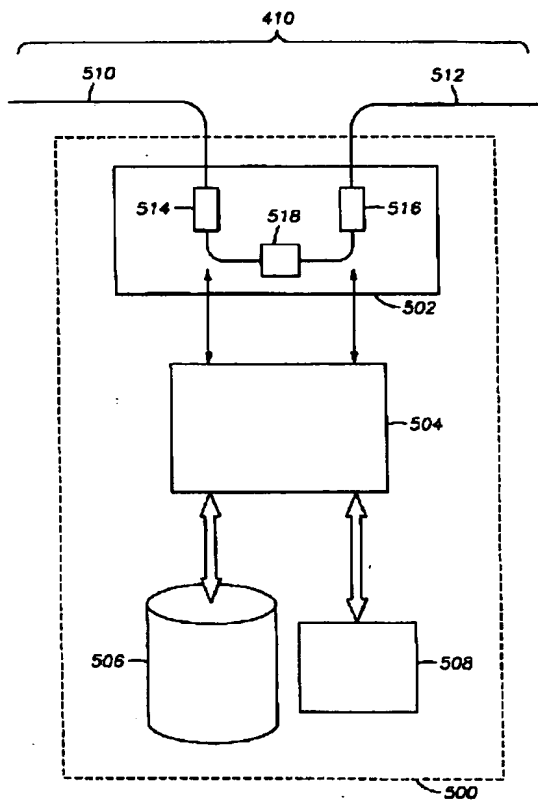
【図 6】



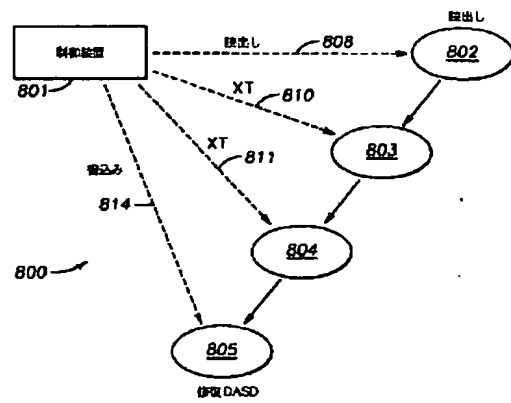
【図 7】



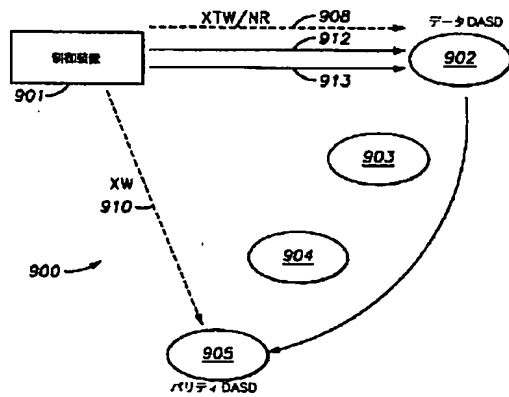
【図 5】



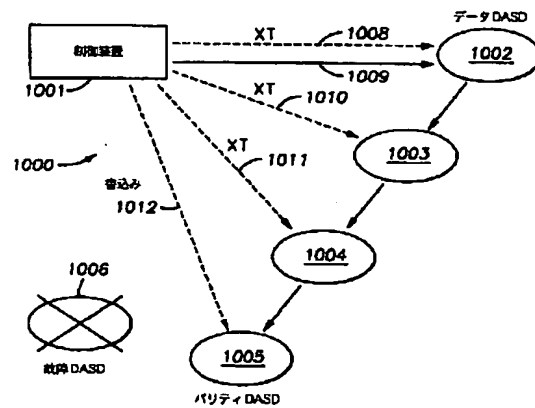
【図 8】



【図 9】



【図 10】



【図11】

